**DOCUMENT RESUME**

ED 325 041 HE 023 959

AUTHOR Tan, David L.
TITLE Is There a Better Way To Measure Quality of Programs?
PUB DATE 11 Sep 90
NOTE 19p.; Paper presented at the Annual Conference of the European Association for Institutional Research (12th, Lyon, France, September 11, 1990).
PUB TYPE Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Departments; *Educational Assessment; *Educational Quality; *Evaluation Methods; Faculty; Higher Education; Input Output Analysis; Institutional Research; *Multivariate Analysis; *Program Effectiveness; *Reputation; School Effectiveness; Student Characteristics

ABSTRACT

The study sought to identify clusters of variables potentially linked to departmental excellence in institutions of higher education, and to determine whether these clusters and combinations thereof could be used to generate rankings of programs which would be consistent with those based on reputation. It studied the interrelationship of variables, within clusters and across clusters. A total of 12 variables, were used: number of faculty, number of graduates produced by the program, number of graduate students, student academic ability, median number of years taken by students to complete their doctorates, placement success rate among graduates in gaining professional employment outside academia, placement success rate among graduates in gaining academic/research positions in doctorate granting universities, library resources, faculty grantsmanship, department research and development spending, program publications, and percentage of faculty members with published articles in a given time period. Principal factor analysis identified three clusters of highly correlated variables, accounting for 67% of the variance. The clusters were labeled faculty research, input cluster, and student cluster. Composite indicators, made up of various combinations of the clusters, were much better than clusters alone at producing ranking etimates similar to those based on reputation. (59 references). (JDD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED325041

# IS THERE A BETTER WAY TO MEASURE QUALITY OF PROGRAMS?

David L. Tan, Ph.D.
Assistant Professor of Higher Education
Department of Educational Leadership
and Policy Studies
College of Education
University of Oklahoma
Norman, Oklahoma 73019-0260
U.S.A.

This paper was presented on September 11, 1990
at the 12th Annual European Association for Institutional
Research Conference in Lyon, France.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

DAVID L. TAN

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

HE 023 957

## IS THERE A BETTER WAY TO MEASURE QUALITY OF PROGRAMS?

*Abstract*

Most quality assessment studies utilized a univariate approach to the assessment of quality. While the univariate approach have some merit, this study presented an alternative--the use of a multivariate approach. Factor analysis identified three clusters of variables potentially linked to quality. These clusters were later used as bases for ranking programs. A total of seven indicators based on these clusters were developed. The indicators which combined the first and second clusters and all three clusters appeared to have produced the best estimates of departmental excellence.

*Introduction*

What is quality in higher education? This perennial question has been and will continue to be a puzzling one. In all likelihood, our obsession with and the search for quality will intensify as higher education moves into the 1990s, a decade predicted to be filled with more demographical, societal, political, and economic factors impacting higher education. Colleges and universities, as in the previous decade, will be unable to isolate themselves from the outside world and will need to be innovative in dealing with all adversities as they come along or face serious threats of financial instability, campus chaos, and much worse--institutional closure.

Historically, higher education institutions have not had much experience managing a steady or declining state. Since the economic recession of the mid-1970s, however, many institutions have had to learn to adapt, sometimes the hard way. Particularly in the late 1970s and early 1980s, many institutions were forced to implement drastic cost measures, such as faculty retrenchment and program reduction just to name a few. While the 1990s may not call for such drastic measures, it could prove to be equally as challenging. Institutions in the 1990s and beyond must continue to discover ways to become more competitive and innovative in all areas of their operations, in their financial strategies, resource reallocation schemes, student recruitment programs, curricular development, and quality enhancement efforts.

Attempts to enhance quality will become particularly critical as issues related to escalating tuition prices, steady state or dwindling financial resources, continuing public pressures for institutional accountability, public disenchantment with higher education, and the perceived decline in the value of higher education begin to take force. The task of enhancing the quality of colleges and universities could have been alleviated if we only know what quality is and how to go about enhancing it in a precise way. Unfortunately, despite much discussion and a myriad of research and scholarly reports published on the subject matter, we are still hard pressed to know the true essence of quality. It remains an elusive concept.

Astin (1985) contends that quality or excellence can be viewed in the context of (a) reputation, (b) resources, (c) outcomes, (d) content, and (e) human talent. Over the years, Astin has been advocating the human talent approach and has felt that it is a better way of viewing institutional excellence. He believes that since student development can essentially be adopted as a goal in all institutions, it makes sense to judge institutions based on their performance on this variable. Other views of excellence (such as those based on reputation, resources, outcomes, and content) can be "counterproductive" in the sense that these approaches do not promote excellence in the higher education system as a whole and do not promote the cause of educational equity. Astin however does realize that the talent development approach has its weaknesses. For example, he has pointed out there is little we know about the causal relationship between financial resources and the talent development of the students. Perhaps, there many other unknowns, such as the effects of varying college environments and institutional types on different forms of student development,

the effects of faculty and student interaction and student campus involvement on affective, cognitive, and moral development of the students, and what specific types of curricula and instruction are better suited at bringing out desired talent development.

While the talent development of students is certainly a desirable institutional outcome and one which is measurable, other variables may be equally as important, such as an institution's performance on its other functions, such as public service and transfer of technology to the private sector and society. While many studies on quality have touched on at least one of the five approaches identified by Astin, there is much overlap. A better approach in sorting through the literature in order to help us understand quality may be to categorize these studies according to the type of research methodology used. One way, as Tan (1986) has done it, was to categorize them into three general types, namely reputational studies, objective indicator studies, and quantitative correlate studies. Briefly, reputational studies are those based on the use of reputation or renown as an indicator of "excellence." Objective indicator studies are those based on so-called "objective" variables, such as resources, outcomes, curricular extensiveness, and the like. And quantitative correlates studies are correlational studies which identify variables highly correlated with excellence. Not surprisingly, the opinions, research methodologies, and findings used and reported in all three types have been as complex and varied as the concept of quality itself. A brief analysis of each type is provided.

*Reputational Studies*

Reputational studies have generally relied on subjective evaluations from faculty, department heads, deans, or panels of "experts" as bases for rating academic programs or institutions. Programs or institutions rated high by consensus are considered highly-reputed. The problem with this type of studies, as many critics have pointed out, is that it is renown and not quality that is being measured and that the two are not necessarily the same thing (Dolan 1976; Lawrence and Green 1980). Further, many reputational studies have been known to suffer from alumni and rater biases. Alumni biases normally occurs when judges are asked to rate programs from which they themselves have graduated (Webster 1981; Cartter 1966; Lawrence and Green 1980). Rater bias normally occurs when evaluators are asked to rate programs unfamiliar to them. In such cases, evaluators simply may have rated programs based on geographical preference or familiarity with the institution rather than the reputation of the specific program in question (Blackburn and Lingenfelter 1973; Lawrence and Green 1980).

*Objective Indicator Studies*

Objective indicator studies, on the other hand, have employed an approach relatively free from subjective judgment. Commonly used indicators have included faculty research productivity, departmental research expenditures, library holdings, student selectivity, student performance, and many others. Conceptually, the use of multiple variables appears to be an appropriate way for measuring a seemingly multidimensional construct like quality but its use has met with limited success. In the absence of an adequate theory of quality, the major problem has been the difficulty in selecting the "best" suited variables or combinations of variables for measuring quality (Smith and Fiedler 1971; Liu 1978). This is perhaps the most frequently cited criticism of objective indicator studies.

*Quantitative Correlate Studies*

The third set of quality assessment studies, quantitative correlate studies, are not designed to measure quality. Rather, their primary purpose are to identify variables that are correlated with excellence, either measured by reputation or some objective indicators. Variables found to have association with excellence are called "quantitative correlates." One of the main criticisms of quantitative correlate studies has been their reliance on reputational ratings. Critics have pointed out that the use of questionable reputational ratings would necessarily yield questionable information about correlates since in almost all cases reputational ratings are used as the dependent variable.

Despite the fact that all three types of studies have their various methodological limitations, they have provided us with some unique paradigms through which we can look at issues and measurement techniques related to the assessment of quality. Many issues however are yet to be resolved. For example, we are still hard pressed to come up with an adequate objective measurement of quality. This present study was an attempt to address this lacuna.

*Purpose of the Study*

There were two purposes of the study. The first purpose was to identify clusters of variables potentially linked to departmental excellence. The second purpose was to determine whether these clusters and combinations thereof could be used to generate rankings of programs and which were better at producing estimates more consistent with those based on reputation.

*Data/Variables Selection*

A total of twelve variables commonly thought of as potentially linked to excellence were used in this study (see Appendix A for a full definition of these variables). The twelve variables included (1) the number of faculty; (2) the number of graduates produced by the program, (3) the number of graduate students enrolled in the program, (4) student academic ability (as measured by the proportion of students who received national fellowships or training support during their graduate education), (5) the median number of years taken by students to complete their doctorates in the program; (6) the placement success rate among graduates in gaining professional employment outside academia, (7) the placement success rate among graduates in gaining academic/research positions in Ph.D.-granting universities, (8) library resources, as measured by a composite library index developed by the Association of Research Libraries, (9) faculty grantsmanship, (10) the amount of departmental research and development spending, (11) the average publications attributed to the program in 1978-80, and (12) the percentage of the faculty members with one or more published articles in 1978-80. These variables were selected for three reasons.

First, data on these variables were readily available in a report compiled by the Conference Board of Associated Councils (Jones et al. 1982). The report covered many disciplines, among them sociology. The second reason for the selection of the twelve variables was that these variables have either been used by researchers in objective indicator studies as indicators of quality or that they have been found to be strong quantitative correlates of excellence. The first variable, the number of faculty members in a department, was a variable either used or found to be a correlate in studies by Saunier (1985), King and Wolfe (1987), the National Science Board (1969), Wispé (1969), Brown (1967), Abbott (1972), Elton and Rodgers (1971), Elton and Rose (1972), Hagstrom (1971), and Oromaner (1970). The second variable, the number of graduates, was a variable used or found as a correlate in studies by Knapp and Goodrich (1952), the National Science Board (1969), Conrad and Blackburn (1985), Elton and Rodgers (1971), Elton and Rose (1972), the Carnegie Council on Policy Studies in Higher Education (1976, 1987), and Tidball and Kristiakowski (1976). The third variable, the number of graduate students enrolled in a program, was used in studies by Elton and Rodgers (1971), Elton and Rose (1972), Conrad and Blackburn (1985), Janes (1969), Lavender et al. (1971) and Morgan et al. (1976). Variables comparable to the fourth variable--the proportion of students who had received national fellowships or training support during their graduate education--were used in studies by Knapp and Greenbaum (1953), Hagstrom (1971), the National Science Board (1969), Gregg and Sims (1972), Astin and Henson (1977), Solmon and Astin (1981), and Conrad and Blackburn (1985). The fifth variable, the median number of years students took to complete their doctorates, was the only variable not previously used or identified as a correlate. The sixth and seventh variables--the two student success rates of gaining employment--were essentially outcomes variables. Such type of variables have appeared in studies by Knapp and Greenbaum (1953), Knapp and Goodrich (1952), Tidball and Kristiakowski (1976), Glower (1980), Dubé (1974) and Brown (1967). The composite library index, the eighth variable, had been used or identified in its various forms in studies by Abbott and Barlow (1972), Brown (1967), Perkins and Snell (1962), Jordan (1963), Conrad and Blackburn (1985), Morgan et al. (1976) and Cartter (1966). Faculty grantsmanship, the ninth

variable, was used or identified in studies by Hagstrom (1971) and Conrad and Blackburn (1985). Departmental research spending, the tenth variable, was used or identified in studies by Glower (1980) and Solmon and Astin (1981). And finally, the eleventh and twelfth variables were related to faculty research productivity. This variable was used or identified in studies by Hagstrom (1971), Guba and Clark (1978), Knudsen and Vaughan (1969), Wispé (1969), Cox and Catt (1977), Abbott (1972), Abbott and Barlow (1972), Drew (1975), Drew and Karpf (1981), Conrad and Blackburn (1985), Somit and Tanenhaus (1964), Lewis (1968), House and Yeager (1978), Glenn and Villimez (1974), Clemente and Sturgis (1974), Liu (1978), Saunier (1985), and King and Wolfe (1987).

The third reason for the selection of these twelve variables was that they were a good representation of the variables that have been used in previous quality assessment studies. Through an extensive literature review, a list of variables used by previous researchers in quality assessment studies was composed. These variables were later sorted and grouped generically into three categories, namely faculty, student, and institutional/departmental (as illustrated in Appendix B). The faculty category consisted of seven generic variables, three of which were represented in this study. The student category consisted of three variables, all of which and an additional variable (namely the length of time students took to complete their doctorate) were represented in this study. Finally, the institutional/departmental category consisted of five variables, two of which were represented in this study.

*Methods of Data Analysis*

Principle component factor analysis, rotated using varimax, was used to identify clusters of highly-correlated variables. The varimax rotation was intended to maximize the variance of the squared loadings in each cluster (i.e. the rotation caused a clearer separation of the clusters). An attempt whenever possible was made to label each cluster according to the type of variables having the heaviest loadings. Assuming that these clusters represented some form of departmental excellence, the relationships of variables within each cluster and across clusters could then be used as bases for ranking programs. To develop departmental rankings, it was necessary to generate a factor score for each significant cluster as well as for all possible combinations of clusters. Each cluster and all possible combination of clusters were called "departmental indicators."

The first departmental indicator was based on the first cluster of highly-correlated variables (the first cluster always explained the most variance in the factor matrix), the second indicator on the second cluster, and so on.

A factor score for each cluster was computed as follows:

> Factor Score = $fsc_{1i}z_1 + fsc_{2i}z_2 + fsc_{ni}z_n$
> (where $fsc_{ji}$ is the factor coefficient for variable j and factor i, and z is the standardized value of the case on variable j) (Nie et al. 1975, p. 489).

In order to develop departmental indicators consisting of multiple clusters, the factor scores from all clusters were combined as follows:

> Composite Score = $x$%*Factor Score (first factor) + $x$%*Factor Score (second factor) ... $x$%*Factor Score (nth. factor)
>
> ($x$ represents the percentage of variance explained by the second factor in the factor structure)

To examine whether these departmental indicators had produced plausible estimates of departmental excellence, they should be compared to a standard, preferably one which was also based on a comparable multivariate approach. Since none as yet existed, the Conference Board's reputational ratings of the faculty were selected. These ratings represented the most recent and best available reputational data because of the much improved methodology used in the collection of the data.

*Assumptions*

This study assumes that quality is a multidimensional construct; accordingly it is susceptible to various interpretations. The definition and measurement of quality in this study was derived through factor analysis of readily available data. This definition and measurement is not claimed to be an absolute for defining and measuring departmental excellence. It is hoped, however, that it can be used to understand the excellence of programs having similar goals and missions. This study, its results, and implications of this study were limited to doctoral sociology programs only.

*Findings*

*Clusters of Departmental Excellence*

Principal factor analysis with varimax rotation identified three clusters of highly-correlated variables, all three accounting for 67 percent of the variance in the factor matrix (Table 1). The other nine clusters were rejected because their eigenvalues were less than 1.0. (A correlation matrix of the twelve variables used in this study is attached as Appendix C.)

**TABLE 1:**
**Rotated Factor Matrix**
(n=92)

| Variable | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|
| nfac | -.08122 | .86515 | .01640 |
| ngrad | .35352 | .69938 | .29023 |
| nstud | .18479 | .85563 | -.06230 |
| sgrant | .64352 | .18213 | .17831 |
| yrphd | -.10925 | .09616 | -.76135 |
| employ1 | .11171 | .10520 | .83660 |
| employ2 | .34207 | .22451 | .76916 |
| library | .72871 | .36049 | -.04368 |
| fgrant | .72947 | .36661 | .05625 |
| money | .17490 | .64276 | .06439 |
| fpub | .76776 | -.05083 | .30212 |
| apub | .86726 | -.05080 | .26113 |

| | Eigenvalue | Percent Variance | Cumulative Variance |
|---|---|---|---|
| Factor 1 | 4.5932 | 38.28 | 35.2 |
| Factor 2 | 2.1243 | 17.70 | 52.9 |
| Factor 3 | 1.3554 | 11.29 | 64.2 |

The first cluster, accounting for 38.3 percent of the variance in the adjusted factor structure, was dominated by variables related to faculty research activity, namely the average faculty publication counts, the percent of faculty with one or more publications, faculty grantsmanship, and library resources. This cluster could be labelled "faculty research."

The second cluster (explanation of 17.7 percent of the factor matrix) was dominated by variables such the number of students enrolled in the program, the number of faculty members, departmental research spending, and the number of graduates produced by the program (Table 1). With the exception of the last variable, all were essentially input variables. Thus, this cluster arguably could be labeled the "input cluster." The fact that this cluster was not dominated by all

input variables may affirm what many researchers and scholars have always thought, which was that quantity (or department size) may not necessarily equal quality. For example, Elton and Rose (1972) in their study of doctoral psychology departments pointed out that "although educational quality and quality need not be mutually exclusive, it may be unwise . . . to conclude that where quantity exists, there also resides quality, or conversely, that is impossible for quality to exist in the absence of large size (p. 200).

The third cluster (explanation of 11.3 percent) appeared to be dominated by variables related to the students. They included the length of time students took to finish their doctorate (yrphd) and how successful they were in finding employment after graduation (employ1 and employ2) (Table 1). This cluster could be labeled the "student" cluster.

*Departmental Indicators*

The three clusters allowed for the construction of a total of seven possible departmental indicators. The first indicator was based on the first cluster, the second indicator on the second cluster, the third indicator on the third cluster, the fourth indicator on the combination of the first and the second clusters, the fifth indicator on the combination of the first and third clusters, the sixth indicator on the combination of the second and the third clusters, and finally, the seventh indicator on the combination of all three clusters. After the seven indicators were developed, factor scores were generated, ranked, and then compared to the Conference Board's reputational rankings.

**TABLE 2:**

**Rank Coefficients**
**Departmental Indicators By Reputation**

| Departmental Indicator # | Combination of Factors | Spearman Rank Coefficient |
|---|---|---|
| 1 | 1 | .677 |
| 2 | 2 | .478 |
| 3 | 3 | .097 |
| 4 | 1 + 2 | .836 |
| 5 | 1 + 3 | .698 |
| 6 | 2 + 3 | .452 |
| 7 | 1 + 2 + 3 | .833 |

From the above table, Departmental Indicators 4 and 7 appeared to have produced more cases of rankings consistent with those based on reputation. A closer examination of the Indicators 4 and 7 was warranted in order to examine how different and/or similar they were in producing rankings. Information on other indicators is also provided below.

**TABLE 3:**
**Rank Comparisons Between**
**Departmental Indicators and Reputation**
(n=54)[1]

| Inst. Name | Reputation | Departmental Indicators 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| Arizona | 9.5 | 10 | 39 | 37 | 14 | 10 | 42 | 16 |
| Boston U. | 31 | 33 | 8 | 50 | 23 | 40 | 27 | 32 |
| Brown | 28.5 | 32 | 21 | 21 | 29 | 30 | 19 | 21 |
| UC Berkeley | 3.5 | 5 | 7 | 40 | 5 | 7 | 13.5 | 6 |
| UC Davis | 31 | 22 | 53 | 43 | 33 | 27 | 53 | 36 |
| UCLA | 9.5 | 23 | 5 | 42 | 12 | 29 | 9 | 17 |
| UC Santa Barbara | 25.5 | 27 | 38 | 13 | 30 | 23 | 31 | 28 |
| Chicago | 1 | 9 | 9 | 7 | 7 | 5 | 7 | 4 |
| Columbia | 7.5 | 7 | 6 | 51 | 4 | 14 | 30 | 9 |
| UConn | 35 | 41 | 24 | 31 | 41 | 42 | 26 | 40 |
| Cornell | 18 | 14 | 27 | 17 | 15 | 12 | 24 | 14 |
| Duke | 25.5 | 4 | 48 | 5 | 10 | 3 | 28 | 8 |
| Florida St. | 40 | 13 | 52 | 41 | 20 | 15 | 50 | 22 |
| Florida | 46 | 18 | 50 | 27 | 26 | 20 | 45 | 27 |
| Georgia | 40 | 17 | 44 | 25 | 22 | 19 | 41 | 24 |
| Harvard | 5.5 | 1 | 23 | 28 | 2 | 1 | 25 | 3 |
| Hawaii | 47.5 | 51 | 40 | 48 | 51 | 52 | 48 | 53 |
| Ill.-Champaign | 18 | 12 | 26 | 16 | 13 | 9 | 21 | 12 |
| Indiana-Bloom. | 12.5 | 26 | 13 | 3 | 19 | 17 | 4 | 13 |
| Iowa St. | 47.5 | 53 | 20 | 1 | 52 | 44 | 3 | 44 |
| Iowa | 37.5 | 20 | 45 | 44 | 27 | 28 | 49 | 33 |
| Kansas | 45 | 40 | 35 | 15 | 46 | 39 | 29 | 41 |
| Louisiana St. | 52.5 | 50 | 43 | 19 | 50 | 45 | 37 | 49 |
| Maryland | 31 | 46 | 19 | 47 | 43 | 48 | 34 | 48 |
| Massachusetts | 20.5 | 36 | 17 | 10 | 36 | 34 | 11 | 31 |
| Michigan St. | 28.5 | 48 | 4 | 49 | 34 | 50 | 10 | 38 |
| Michigan | 3.5 | 11 | 2 | 23 | 1 | 11 | 2 | 2 |
| Minnesota | 23 | 21 | 11 | 26 | 17 | 21 | 13.5 | 18 |
| Missouri | 35 | 31 | 33 | 34 | 35 | 33 | 36 | 35 |
| Nebraska | 43 | 34 | 49 | 18 | 42 | 32 | 39 | 39 |
| New York U. | 23 | 49 | 18 | 29 | 47 | 47 | 22 | 46 |
| N. Carolina | 5.5 | 6 | 32 | 4 | 9 | 4 | 16 | 7 |
| Northwestern | 23 | 24 | 34 | 33 | 24 | 25 | 35 | 26 |
| Ohio St. | 35 | 28 | 28 | 12 | 31 | 26 | 20 | 29 |
| Oklahoma St. | 52.5 | 52 | 25 | 9 | 53 | 49 | 15 | 52 |
| Pennsylvania | 12.5 | 29 | 15 | 30 | 25 | 31 | 17 | 25 |
| Pittsburgh | 33 | 42 | 22 | 39 | 40 | 43 | 32 | 43 |
| Princeton | 16 | 16 | 42 | 35 | 21 | 18 | 43 | 23 |
| Purdue | 37.5 | 39 | 37 | 6 | 45 | 35 | 18 | 37 |
| Rutgers | 27 | 47 | 3 | 38 | 28 | 46 | 6 | 30 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| SUNY Stony B. | 14.5 | 43 | 12 | 11 | 38 | 37 | 8 | 34 |
| So. Calif. | 20.5 | 3 | 36 | 46 | 8 | 6 | 47 | 11 |
| So. Illinois | 49.5 | 37 | 41 | 32 | 44 | 41 | 40 | 45 |
| Stanford | 7.5 | 2 | 46 | 22 | 6 | 2 | 38 | 5 |
| Syracuse | 40 | 30 | 47 | 45 | 37 | 36 | 51 | 42 |
| Temple | 43 | 44 | 14 | 53 | 39 | 53 | 46 | 50 |
| Tennessee | 49.5 | 38 | 51 | 20 | 49 | 38 | 44 | 47 |
| TX.-Austin | 18 | 25 | 10 | 24 | 18 | 22 | 12 | 19 |
| Wash. U. (MO) | 43 | 35 | 16 | 2 | 32 | 24 | 5 | 20 |
| Washington | 11 | 8 | 30 | 36 | 11 | 8 | 33 | 10 |
| Wayne St. | 51 | 45 | 31 | 52 | 48 | 51 | 52 | 51 |
| Wisconsin | 2 | 19 | 1 | 8 | 3 | 16 | 1 | 1 |
| Yale | 14.5 | 15 | 29 | 14 | 16 | 13 | 23 | 15 |

[1] 39 programs were not ranked because they had missing data on one or more variables.

As can be examined in the above table, programs which were quite similarly ranked by Departmental Indicators 4 and 7 and by reputation included those at the University of California-Berkeley (ranked about the fifth position by the three systems), the University of Michigan (ranked in the top third by the three systems), the University of Missouri (identically ranked at thirty-fifth by the three systems), Northwestern University (ranked about twenty-fourth), Oklahoma State University (ranked about fifty-second), the University of Texas-Austin (ranked about eighteenth), the University of Washington (ranked around tenth), the University of Wisconsin (ranked in the top third), and Yale University (ranked about the fifteenth position by the three systems).

There were also cases in which the multivariate approaches had produced drastically different results compared to reputational rankings. Such examples include Duke University, Florida State University, Florida, Georgia, Maryland, New York University, SUNY Stony Brook, and Washington University (Missouri). In the case of Duke University, the sociology program there was ranked about the tenth position by the Departmental Indicators 4 and 7 but was ranked twenty-fifth by reputation. Duke's program should have been ranked higher because it performed very well on the first factor (a rank of fourth) and had achieved a rank of fifth on the third factor. It did not perform well however on the second factor--it had achieved a rank of forty-eighth. Consequently, Indicator 7 ranked it about two positions higher than did Indicator 4.

The two most visible programs in which there were large differences in rankings between the multivariate approaches (Indicators 4 and 7) and reputation were the sociology programs at New York University and Washington University. The program at New York University was ranked much lower than its reputational rank of twenty-third and justifiably so because it did not perform well at all on factor 1 (a rank of only forty-ninth). On the other hand, the program at Washington University should have been ranked considerably higher than its reputational rank of forty-third because it performed quite well on the second factor (a rank of 16) and the third factor (a rank of second).

There were also instances in which Departmental Indicators 4 and 7 had produced quite different ranks from each other. For example, the program at Temple University was ranked much lower by Indicator 7 than by Indicator 4--a rank of 50 compared to 39. Indicator 4 did not take into consideration factor 3, a dimension in which this program did not perform particularly well--it received a low rank of 53 for its performance on this factor. As a result, Indicator 4 rated this program much higher than it deserved. Another example of such discrepancy occurred for the program at Washington University. Indicator 7 rated the program much higher than Indicator 4 for the reason that it performed well on the third factor. The rank produced by Indicator 4 however was closer to the program's reputational ranking.

In sum, it appeared that the composite indicators, especially Departmental Indicators 4 and 7, were much better at producing ranking estimates similar to those based on reputation. The best possible indicator was the one based on the first and second factors while the second best (only by a small margin) was based on all three factors. It is conceivable that these two indicators may even be better than reputational ratings because they do take into account the complex interrelationships of variables potentially linked to quality.

*Discussion*

While the univariate approach to quality assessment deserves some merits, especially in the absence of a theory of quality, the use of the multivariate approach would seem to be a better approach since quality seems likely to be a multidimensional construct. This study has shown that the interrelationships of variables potentially linked to quality can be examined and studied in some depth, unlike past methodologies which have not allow it to be possible. How different is this methodology from past quality assessment studies?

In reputational studies, the focus has mainly been on the ratings of programs based on reputation (Cartter 1966; Roose and Andersen 1970; Magoun 1966; National Science Board 1969; Petrowski et al. 1973; Jones et al. 1982; Margulies and Blau 1973; Blau and Margulies 1974; and Cartter and Solmon 1977). Therefore, the interrelationship of many other important variables related to quality were rarely examined and almost never incorporated in the computation of a program rank.

Many objective indicator studies also suffer from this limitation. The single most often used variable for rating programs appears to be faculty research productivity (Somit and Tanenhaus 1964; Lewis 1968; Knudsen and Vaughan 1969; Glenn and Villimez 1970; Clemente and Sturgis 1974; Wispé 1969; Cox and Catt 1977; Liu 1978; Gill 1975; and House and Yeager 1978). Obviously, faculty research productivity does not equal faculty quality nor does it equal program quality. Other studies have used student career success (Glower 1980), academic ability of students (Astin and Henson 1977), the number of alumni who went on and earned doctoral degrees (Knapp and Goodrich 1952; Knapp and Greenbaum 1953; Tidball and Kristiakowski 1976), student research productivity (Dubé 1974), college preferences of highly-able students (Astin and Solmon 1979), and the number of degrees awarded (Carnegie Council on Policy Studies in Higher Education 1976). The question that still remains is which of the above variables can stand as an adequate surrogate indicator of quality.

Not all objective indicator studies, however, used the univariate approach. The two notable exceptions were the controversial Gourman reports (1967-1983) and the study by Brown (1967). The Gourman reports (1967-1983) claimed to have employed a multivariate approach in the ranking of a variety of programs, but many scholars do not believe it (Webster 1984). The methodology in the Brown study was more convincing but unfortunately its primary focus was on ranking institutions rather than on studying the interrelationship of variables potentially linked to quality.

Quantitative correlate studies, by comparison, have examined the interrelationship of variables more adequately. The Hagstrom (1971) study, perhaps the best known of the quantitative correlate studies, examined the correlates of quality through a series of product-moment correlations and multiple linear regression analyses. The study found, for example, a strong interaction between faculty research productivity, faculty grantsmanship, and the number of faculty in a program and that these variables were major correlates of reputation. The major limitation of this study, like many others before it, was that it still had to be reliant on reputational ratings as the dependent variable. In the Hagstrom study, reputational ratings compiled by Cartter (1966) were used. This limitation, however, did not apply to the quantitative correlate study undertaken by Conrad and Blackburn (1985). These authors used program reviews instead of reputational ratings as the dependent variable and focused their study on regional colleges and universities instead of doctoral-granting universities. One of the major findings of the study was that faculty research productivity, faculty grantsmanship, the number of students, the average academic ability of students, the size of the library, and curricular concentration were the best correlates of program excellence. Some of the findings obtained by Hagstrom (1971) and Conrad

and Blackburn (1985) were confirmed in this present study. But the differences are what made this study unique.

First, this study was able to reveal the interrelationship of variables in the form of a mathematical factor matrix. The methodology used in many quantitative correlate studies, by comparison, was only appropriate at identifying the correlates of quality (usually reputation) and how these correlates interacted with each other in terms of their explanations of variance in the dependent variable. No combination of such correlates was ever attempted. Even if such correlates could be combined, the end results would serve little purpose since the composite variable would be similar to the reputational ratings (or program reviews) used as the dependent variable in these studies.

Second, this study represented an attempt to resolve many of the inadequacies reflected in previous studies. The major strengths were that it permitted the assessment of departmental excellence to be relatively free from subjective evaluation and also allowed for the study of the interrelationship of variables, not just within each cluster of highly-correlated variables but across clusters as well. To be sure, the biggest advantage of the study was its multivariate approach. A ranking system based on such an approach would not be as affected by fluctuations in just one or two variables, a problem common to many univariate approaches.

### *Concluding Remarks*

This study was a start toward the use of the multivariate approach, but whether or not the clusters identified in the study were adequate indicators of quality remain yet to be seen. One could argue for the inclusion of additional variables to the factor equation. Regardless of how many other additional variables are added, the chicken-egg problem associated with the lack of an adequate theory of quality will remain a difficult and puzzling one to solve, but it is a problem which should not deter those of us who need to know about quality and how we can make better decisions about enhancing the quality of our academic programs and our institutions.

There remains yet at least two other issues. The first is that even if we are able to come up with an adequate definition or measurement of quality, should that definition or measurement necessary apply to all institutions of higher education? Given the diversity of higher education institutions in this country and even some programs in the same discipline, should the same definition necessarily apply to all? Should a model of quality based on characteristics prevalent at elite institutions be applicable to non-elite (or "representative") institutions, and vice versa?

The second issue has to do with the relationship between quality and the value-added development of students. As Astin and Solmon (1981, p. 19) put it clearly: "Do students in highly rated . . . institutions develop differently from students in institutions with mediocre or low ratings?" This "value-added" question may be difficult to answer until we resolve the issue of quality first. In other words, we must first identify which of our institutions or programs are more competitive and which are less competitive (based on criteria other than student development), and then conduct a longitudinal study to determine whether students develop differently in different types of institutions.

In conclusion, it is evident that more knowledge is still needed about quality. This study has shed some light on the topic but it is a only a beginning. Perhaps future research could be focused on other disciplines to determine if the same clusters appear. The relationships of variables within and across clusters could be examined more closely in order that we may understand better how different types of variables interact with each other in other disciplines and how all affect the overall picture of what we are trying to understand.

## APPENDIX A:
## Variables Used in the Study

| Variable Name | Description of Variable |
|---|---|
| Number of Faculty (nfac) | Reported number of faculty members in the program, December 1980. |
| Number of Graduates (ngrad) | Reported number of program graduates from July 1975 to June 1980. |
| Number of Students (nstud) | Reported total number of full-time and part-time graduate students enrolled in the program who intended to earn doctorates, December 1980. |
| Student Fellowship Awards (sgrant) | The proportion of FY 1975-79 program graduates who had received national fellowships or training support during their graduate education. |
| Years Taken to Complete Doctorate (yrphd) | The median number of years from first enrollment in graduate school to receipt of the doctorate among FY 1975-79 program graduates. |
| Student Placement Success Rate 1 (employ1) | The proportion of program graduates who at the time they completed their doctorate reported that they had been offered professional employment outside academia (data were collected from FY 1975 through 1979). |
| Student Placement Success Rate 2 (employ2) | The proportion of program graduates who at the time they completed their doctorate reported that they had been offered academic positions in Ph.D.-granting universities (data were collected from FY 1975 through 1979). |
| Library Resources (library) | Composite index describing the library in the university in which the program is located, 1979-80. (This index was developed by the Association of Research Libraries.) |
| Faculty Grantsmanship (fgrant) | The proportion of faculty members in the program holding research grants from the three federal agencies (Alcohol, Drug Abuse, and Mental Health Administration, National Institutes of Health, or the National Science Foundation) at any time during FY 1978-1980. |
| Departmental Expenditures (money) | The total expenditures reported by the university for research and development activities in the specified field in FY 1979. |
| Faculty Research Productivity (fpub) | The average number of publication articles attributed to the faculty in the program from 1978-79. This variable was derived by dividing the total number of faculty publications by the number of faculty in the program (nfac). |

| | |
|---|---|
| Average Publication Records (apub) | The percent of faculty members with one or more published articles in 1978-90. |

Source: Jones, Lyle V., Lindzey Gardner, and Porter E. Coggeshall (eds.). *An Assessment of Research-Doctorate Programs in the United States.* Washington, D.C.: National Academy Press, 1982.

## APPENDIX B

### Variables Used as Objective Indicators of Quality or Those Identified as Quantitative Correlates of Quality

*Faculty*

1. Faculty research productivity (Hagstrom 1971; Guba and Clark 1978; Knudsen and Vaughan 1969; Wispé 1969; Abbott 1972; Abbott and Barlow 1972; Jones et al. 1982; Drew 1975; Cox and Catt 1977; Drew and Karpf 1981; Conrad and Blackburn 1985; Somit and Tanenhaus 1964; Lewis 1968; House and Yeager 1978; Glenn and Villimez 1974; Clemente and Sturgis 1974; Liu 1978); Saunier (1985); King and Wolfe (1987).

2. Faculty grantsmanship (Hagstrom 1971; Jones et al. 1982; Conrad and Blackburn 1985).

3. Number of faculty (National Science Board 1969; Abbott 1972; Wispé 1969; Brown 1967; Abbott 1972; Elton and Rodgers 1971; Elton and Rose 1972; Jones et al. 1982; Oromaner 1970; Saunier 1985; King and Wolfe 1987).

4. Faculty background (sex, age at Ph.D. receipt, professional age, academic origins, etc.) (Clemente and Sturgis 1974; Oromaner 1970; Conrad and Blackburn 1985; Brown 1967; National Science Board 1969; Abbott and Barlow 1972; Axelson 1960).

5. Faculty honorarium (awards, journal editorship, society membership) (Hagstrom 1971; Bowker 1965).

6. Faculty workload (Conrad and Blackburn 1985).

7. Faculty compensation/fringe benefits (National Science Board 1969; Abbott and Barlow 1972; Brown 1967).

*Student*

1. Number of graduate students (Elton and Rodgers 1971; Elton and Rose 1972; Conrad and Blackburn 1985; Janes 1969; Lavender et al. 1971; Morgan et al. 1976).

2. Academic ability of students (Knapp and Greenbaum 1953; Hagstrom 1971; National Science Board 1969; Gregg and Sims 1972; Astin and Henson 1977; Solmon and Astin 1981; Conrad and Blackburn 1985).

3. Outcomes (number of degrees awarded; proportion of students who went on and earned doctoral degrees; etc.) (Knapp and Greenbaum 1953; Knapp and Goodrich 1952; Tidball and Kristiakowski 1976; National Science Board 1969; Conrad and Blackburn 1985; Elton and Rodgers 1971; Elton and Rose 1972; Dubé 1974; Brown 1967; Carnegie Council on Policy Studies in Higher Education 1976, 1987; Glower 1980; Jones et al. 1982).

*Institutional/Departmental Variables*

1. Curricular concentration/specialization (Conrad and Blackburn 1985; Elton and Rodgers 1971; Elton and Rose 1972; Solmon and Astin 1981).

2. Federal funding (National Science Board 1969; Carnegie Council on Policy Studies 1976, 1987).

3. Library resources (Abbott and Barlow 1972; Perkins and Snell 1962; Brown 1967; Jordan 1963; Morgan et al. 1976; Conrad and Blackburn 1985; Cartter 1966; Jones et al. 1982).

4. Physical resources (Abbott and Barlow 1972).

5. Departmental research spending (Jones et al. 1982; Kelso 1975; Glower 1980; Solmon and Astin 1981).

## APPENDIX C:
## Correlation Matrix[1]

| | nfac | grad | stud | sgrant | yrphd | employ1 | employ2 | library | fgrant | money | fpub | apub |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| nfac | 1.000 | | | | | | | | | | | |
| grad | .552 | 1.000 | | | | | | | | | | |
| stud | .625 | .675 | 1.000 | | | | | | | | | |
| sgrant | .044 | .310 | .148 | 1.000 | | | | | | | | |
| yrphd | -.203 | -.179 | .086 | -.226 | 1.000 | | | | | | | |
| employ1 | .061 | .186 | -.003 | .292 | -.293 | 1.000 | | | | | | |
| employ2 | .239 | .316 | .115 | .388 | -.415 | .503 | 1.000 | | | | | |
| library | .435 | .545 | .467 | .205 | -.105 | .070 | .343 | 1.000 | | | | |
| fgrant | .179 | .463 | .434 | .431 | -.156 | .143 | .358 | .567 | 1.000 | | | |
| money | .372 | .375 | .295 | .312 | -.101 | .149 | .222 | .258 | .383 | 1.000 | | |
| fpub | .167 | .254 | .043 | .345 | -.495 | .159 | .381 | .481 | .346 | .107 | 1.000 | |
| apub | .078 | .352 | .098 | .430 | -.355 | .226 | .388 | .528 | .554 | .199 | .724 | 1.000 |

[1]92 Total Cases
Of which 39 are missing in at least in one variable
Pairwise Deletion of Missing Cases

*References*

Abbott, W.F.(1972). "University and Departmental Determinants of the Prestige of Sociology Departments." *American Sociologist*, 7 (November): 14-15.

Abbott, W.F. and H.M. Barlow (1972). "Stratification Theory and Organizational Rank: Resources, Functions and University Prestige in the United States." *Pacific Sociological Review*, 15 (October): 401-424.

Astin, A.W. (1985). *Achieving Educational Excellence.* San Francisco: Jossey-Bass,

Astin, A.W. and J.W. Henson (1977). "New Measures of College Selectivity." *Research in Higher Education*, 6: 1-8.

Astin, W. and L.C. Solmon (1981). "Are Reputational Ratings Needed to Measure Quality?" *Change*, 13 (October): 14-19.

Axelson, L.J. (1960). "Graduate Schools and the Productivity of Their Graduates." *American Journal of Sociology*, 66 (September): 171-175.

Blackburn, R.T. and P.E. Lingenfelter (1973). *Assessing Quality in Doctoral Programs: Criteria and Correlates of Excellence.* Ann Arbor, Michigan: University of Michigan.

Bowker, A.H. (1965). "Quality and Quantity in Higher Education." *Journal of the American Statistical Association*, 60 (March): 1-15.

Brown, D.G. (1967). *The Mobile Professors.* Washington, D.C.: The American Council on Education.

Carnegie Council on Policy Studies in Higher Education (1976). *A Classification of Institutions of Higher Education.* Berkeley, California: The Carnegie Foundation for the Advancement of Teaching.

---------- (1987). *A Classification of Institutions of Higher Education.* Princeton, N.J.: The Carnegie Foundation for the Advancement of Teaching.

Cartter, A.M. (1966). *An Assessment of Quality in Graduate Education.* Washington, D.C.: The American Council on Education.

Clemente, F. and R. Sturgis (1974). "Quality of Department of Doctoral Training on Research Productivity." *Sociology of Education*, 47: 287-299.

Conrad, C.F. and R.T. Blackburn (1985). "Correlates of Departmental Quality in Regional Colleges and Universities." *American Educational Research Journal*, 22 (Summer): 279-295.

Cox, W. M. and V.Catt (1977). "Productivity Ratings of Graduate Programs in Psychology Based on Publication in the Journals of the American Psychological Association." *American Psychologist*, 32: 793-813.

Dolan, W. P. (1976). *The Ranking Game: The Power of the Academic Elite.* Lincoln, Nebraska: The University of Nebraska Printing and Duplicating Service.

Drew, D.E. (1975). *Science Development: An Evaluation Study.* Washington, D.C.: The National Academy of Sciences.

Drew, D.E. and R.S. Karpf (1981). "Ranking Academic Departments: Empirical Findings and a Theoretical Perspective." *Research in Higher Education,* 14: 305-320.

Dubé, W.F. (1974). "Undergraduate Origins of U.S. Medical Students." *Journal of Medical Education,* 49: 1005-1010.

Elton, C.F. and S.A. Rodgers (1971). "Physics Department Ratings: Another Evaluation." *Science,* 174 (November): 565-568.

Elton, C.F. and Harriett A. Rose (1972). "What Are the Ratings Rating?" *American Psychologist,* 27: 197-201.

Glenn, N.D. and W.Villimez (1970). "The Productivity of Sociologists at 45 American Universities." *American Sociologist,* 5 (August): 244-252.

Glower, D.D. (1980). "A Rational Method for Ranking Engineering Programs." *Engineering Education,* 70: 788-794; 842.

Gourman, J. (1967). *The Gourman Report: Ratings of American Colleges.* Phoenix, Arizona: The Continuing Research Institute, 1967.

---------- (1977a). *The Gourman Report: A Rating of Undergraduate Programs in American and International Universities.* Los Angeles: The National Education Standards.

---------- (1977b). *The Gourman Report: A Rating of Graduate and Professional Programs in American and International Universities.* Los Angeles: The National Education Standards.

---------- (1982). *The Gourman Report: A Rating of Undergraduate Programs in American and International Universities.* Los Angeles: The National Education Standards.

---------- (1983). *The Gourman Report: A Rating of Graduate and Professional Programs in American and International Universities.* Revised Second Edition. Los Angeles: The National Education Standards.

Gregg, R.T. and P.D. Sims (1982). "Quality of Faculties and Programs of Graduate Departments of Educational Administration." *Educational Administration Quarterly,* 8: 67-92.

Gross, G.R. (1970). "The Organizational Set: A Study of Sociology Departments." *American Sociologist,* 5: 25-29.

Guba, E.G. and D.L. Clark (1978). "Levels of R & D Productivity in Schools of Education." *Educational Researchers,* 7: 3-9.

Hagstrom, W.O. (1971). "Inputs, Outputs, and the Prestige of University Science Departments." *Sociology of Education,* 44 (Fall): 375-397.

House, D.R. and J.H. Yeager, Jr. (1978). "The Distribution of Publication Success Within and Among Top Economic Departments: A Disaggregate View of Recent Evidence." *Economic Inquiry,* 16 (October): 593-598.

Janes, R.W. (1969). "The Student Faculty Ratio in Graduate Programs of Selected Departments of Sociology." *American Sociologist*, 4: 123-127.

Jones, Lyle V., G. Lindzey, and P.E. Coggeshall (eds.). (1982). *An Assessment of Research-Doctorate Programs in the United States: Social and Behavioral Sciences.* Washington, D.C.: The National Academy Press.

Jordan, R.T. (1963). "Library Characteristics of Colleges Ranking High in Academic Excellence." *College and Research Libraries,* 24: 369-376.

Kelso, C.D. (1975). "How Does Your Law School Measure Up?" *Student Lawyer*, 4: 20-24.

King, S. and L.M. Wolfe (1987). "A Latent-Variable Causal Model of Faculty Reputational Ratings." *Research in Higher Education,* 27: 99-106.

Knapp, R.H. and J.J. Greenbaum (1953). *The Younger American Scholar: His Collegiate Origins.* Chicago: The University of Chicago Press.

Knapp, R.H.and H.B. Goodrich (1952). *Origins of American Scientists.* New York: Russell and Russell.

Knudsen, D.D. and T.R. Vaughan (1969). "Quality in Graduate Education: A Re-evaluation of the Rankings of Sociology Departments in the Cartter Report." *American Sociologist,* 4 (February): 12-19.

Lavender, A.D., R.A. Mathers, and J. Pease (1971). "The Student-faculty Ratio in Graduate Programs of Selected Departments of Sociology: A Supplement to the Janes Report." *American Sociologist,* 6: 29-30.

Lawrence, J.K. and K.C. Green (1980). *A Question of Quality: The Higher Education Ratings Game.* AAHE-ERIC/Higher Education Research Report No. 5. Washington, D.C.: American Association for Higher Education.

Lewis, L.S. (1968). "On Subjective and Objective Rankings of Sociology Departments." *American Sociologist,* 3 (May): 129-131.

Liu, H.C. (1978). "Faculty Citation and Quality of Graduate Engineering Departments." *Engineering Education,* 68 (April): 739-741.

Morgan, D.L., R.C. Kearney, and J.L. Regens (1976). "Assessing Quality Among Graduate Institutions of Higher Education in the United States." *Social Science Quarterly,* 57 (December): 670-679.

National Science Board (1969). *Graduate Education: Parameters For Public Policy.* Washington, D.C.: The National Science Foundation.

Nie, N.H., C. H. Hull, J.G. Jenkins, K. Steinbrener, and D.H. Bent 1975). *SPSS: Statistical Package for the Social Sciences.* Second Edition. New York: McGraw-Hill Book Company.

Oromaner, M.J. (1970). "A Note on Analytical Properties and Prestige of Sociology Departments." *American Sociologist,* 5 (August): 240-244.

Perkins, D. and J.L. Snell (1962). *The Education of Historians in the United States.* New York: McGraw-Hill.

Saunier, M.E. (1985). "Objective Measures as Predictors of Reputational Ratings." *Research in Higher Education,* 23: 227-244.

Smith, R. and F.E. Fiedler (1971). "The Measurement of Scholarly Work: A Critical Review of the Literature." *Educational Record,* 52 (Summer): 225-232.

Solmon, L. and A.W. Astin (1981). "Departments Without Distinguished Graduate Programs." *Change,* 13 (October): 23-28.

Somit, A. and J.Tanenhaus (1964). *American Political Science: A Profile of a Discipline.* New York: Atherton Press.

Tan, D.L. (1986) "The Assessment of Quality in Higher Education: A Critical Review of the Literature and Research." *Research in Higher Education,* 24 (3): 223-265.

Tidball, M.E. and V. Kristiakowski (1976). "Perspective on Academic Women and Affirmative Action." *Educational Record,* 54: 130-135.

Webster, D.S. (1981). "Methods of Assessing Quality." *Change,* 13 (October): 20-24.

---------- (1984). "Who is Jack Gourman and Why Is He Saying All These Things About My College?" *Change* (Nov./Dec): 14-19.

Wispé, L.G. (1969). "The Bigger the Better: Productivity, Size, and Turnover in a Sample of Psychology Departments." *American Psychologist,* 24: 662-668.